

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Differentially methylated microRNAs in prediagnostic samples of subjects who developed breast cancer in the european prospective investigation into nutrition and cancer (EPIC-Italy) cohort**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1532163> since 2022-06-24T12:58:10Z

*Published version:*

DOI:10.1093/carcin/bgv102

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

ORIGINAL MANUSCRIPT

# Differentially methylated microRNAs in prediagnostic samples of subjects who developed breast cancer in the European Prospective Investigation into Nutrition and Cancer (EPIC-Italy) cohort

Francesca Cordero<sup>1</sup>, Giulio Ferrero<sup>1,2</sup>, Silvia Polidoro<sup>3</sup>, Giovanni Fiorito<sup>4,5</sup>, Gianluca Campanella<sup>6</sup>, Carlotta Sacerdote<sup>7</sup>, Amalia Mattiello<sup>8</sup>, Giovanna Masala<sup>9</sup>, Claudia Agnoli<sup>10</sup>, Graziella Frasca<sup>11</sup>, Salvatore Panico<sup>8</sup>, Domenico Palli<sup>9</sup>, Vittorio Krogh<sup>10</sup>, Rosario Tumino<sup>11</sup>, Paolo Vineis<sup>3,12</sup> and Alessio Naccarati<sup>3,\*</sup>

<sup>1</sup>Department of Computer Science, University of Torino, Corso Svizzera 185, 10149 Turin, Italy, <sup>2</sup>Department of Clinical and Biological Sciences, University of Torino, 10043 Orbassano, Italy, <sup>3</sup>Molecular and Genetic Epidemiology Unit and <sup>4</sup>Genomic Variation in Human Populations and Complex Diseases Unit, Human Genetics Foundation, Via Nizza 52, 10126 Turin, Italy, <sup>5</sup>Department of Clinical and Biological Sciences, University of Torino, Via Santena 7, 10126 Turin, Italy, <sup>6</sup>Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus Norfolk Place W2 1PG London, UK, <sup>7</sup>Unit of Cancer Epidemiology, AO Citta' della Salute e della Scienza-University of Torino and Center for Cancer Prevention (CPO), Via Santena 7, 10126 Turin, Italy, <sup>8</sup>Department of Clinical Medicine and Surgery, Federico II University, Via S. Pansini 5, 80131 Naples, Italy, <sup>9</sup>Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute - ISPO, Via delle Oblate 4, 50141 Florence, Italy, <sup>10</sup>Epidemiology and Prevention Unit Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian 1, 20133 Milan, Italy, <sup>11</sup>Epidemiology and Prevention Unit Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian 1, 20133 Milan, Italy, <sup>12</sup>MRC-PHE Centre for Environment and Health, Imperial College London, St Mary's Campus Norfolk Place W2 1PG London, UK,

\*To whom correspondence should be addressed. Human Genetics Foundation Torino (HuGeF), Via Nizza, 52 Torino, Italy. Tel: +39 0116709539; Fax: +39 0116709541; Email: [alessio.naccarati@hugef-torino.org](mailto:alessio.naccarati@hugef-torino.org)

## Abstract

The crosstalk between microRNAs (miRNAs) and other epigenetic factors may lead to novel hypotheses about carcinogenesis identifying new targets for research. Because a single miRNA can regulate multiple downstream target genes, its altered expression may potentially be a sensitive biomarker to detect early malignant transformation and improve diagnosis and prognosis. In the current study, we tested the hypothesis that altered methylation of miRNA encoding genes, associated with deregulated mature miRNA expression, may be related to dietary and lifestyle factors and may contribute to cancer development. In a case-control study nested in a prospective cohort (EPIC-Italy), we analysed DNA methylation levels of miRNA encoding genes (2191 CpG probes related to 517 genes) that are present in the Infinium Human Methylation450 BeadChip array in prediagnostic peripheral white blood cells of subjects who developed colorectal cancer (CRC,  $n = 159$ ) or breast cancer (BC,  $n = 166$ ) and matched subjects who remained clinically healthy. In the whole cohort, several differentially methylated miRNA genes were observed in association with age, sex, smoking habits and physical activity. Interestingly, in the case-control study, eight differentially methylated miRNAs were identified in subjects who went on to develop BC (miR-328, miR-675, miR-1307, miR-1286, miR-1275, miR-1910, miR-24-1 and miR-548a-1; all Bonferroni-adjusted  $P < 0.05$ ). No significant associations were found with CRC. Assuming that altered methylation of miRNAs detectable in blood may be present before diagnosis, it may represent a biomarker for early detection or risk of cancer and may help to understand the cascade of events preceding tumour onset.

Received: February 6, 2015; Revised: June 30, 2015; Accepted: July 8, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

## Abbreviations

BC	breast cancer
BMI	body mass index
CRC	colorectal cancer
CGI	CpG islands
miRNAs	microRNAs
RPM	recursively partitioned mixture model
SNPs	single-nucleotide polymorphisms
TSS	transcription start site

## Introduction

DNA methylation is an epigenetic mechanism responsible for heritable phenotypic changes of cells with functional consequences. It is an important phenomenon in early embryonic development, stem cell differentiation and tissue-specific gene expression (1). Methyltransferases add methyl groups to cytosine residues followed by guanines (CpGs). CpG islands (CGI) are CpG rich regions often found in gene promoters, and the methylation state of a CpG island may correlate with the gene expression state (2). In normal cells, promoter-associated CGI are generally unmethylated, with the exception of X-chromosome inactivation or genomic imprinting. In cancer, tumour cells are characterized by global hypomethylation, but locus-specific hypermethylation of promoter-associated CGI resulting in gene silencing (3).

MicroRNAs (miRNAs) are short non-coding RNA molecules (18–25 nucleotides) able to repress the translation of multiple protein-coding mRNAs by sequence-specific binding to the 3' untranslated region (4). It is currently estimated that there are more than 1800 mature miRNAs encoded from thousands of transcriptional units (miRNA genes) in the human genome ([www.mirBase.org](http://www.mirBase.org)). Based on their genomic location, miRNA genes can be classified as intergenic or intragenic. Intergenic miRNAs are transcribed from non-coding regions in-between protein-coding and other non-coding genes. Instead intragenic miRNAs are encoded in the intron of, and mostly transcribed in parallel with, their host protein-coding or non-coding gene (5). MiRNAs participate in the regulation of about two-thirds of human genes and are involved in the determination of cell identity. The rapid developments in miRNA-related technologies, such as miRNA expression profiling and synthetic oligoRNA, have contributed to identify miRNAs involved in a number of physiological and pathological phenotypes (6). Even though the biogenesis of miRNAs has been intensely studied and described, the regulation of miRNA expression is not fully understood. Recently, accumulating studies have shown that subgroups of miRNAs are epigenetically regulated (7). MiRNA-coding genes appear to be regulated in a fashion similar to protein-coding genes, through the actions of transcription factors and epigenetic control mechanisms such as methylation of specific regulatory regions (7,8). As a consequence, alteration of these regulatory mechanisms can produce abnormal chromatin states and participate in disease pathogenesis. For example, aberrant methylation of tumour suppressive miRNAs has been reported in different cancer types (9).

For these reasons, miRNA targets of aberrant DNA methylation might be potentially used for diagnostic purposes, but also as potential targets for epigenetic drugs. Conversely, other subsets of miRNAs may control the expression of important epigenetic regulators, including DNA methyltransferases, histone deacetylases and Polycomb-group genes. This complicated network of feedback between miRNAs and epigenetic pathways

appears to form regulatory circuits and organize the whole gene expression profile (6). However, several questions remain largely unanswered, such as: (i) for how many miRNAs the expression is effectively controlled by methylation (or in concomitance with other epigenetic mechanisms), (ii) what are the consequences of an aberrant miRNA methylation and (iii) which genes are regulated by each miRNA.

The analysis of the crosstalk between miRNAs and other epigenetic modifications may lead to novel hypotheses about cancer onset and may identify new targets for research and hopefully therapies. There are several studies reporting deregulated methylation patterns in miRNAs, mostly based on a candidate gene approach. Shen *et al.* previously reported results from an analysis of 62 paired tumour and adjacent non-tumour tissues from hepatocellular carcinoma cases using the Infinium HumanMethylation27 DNA Analysis BeadChip, identifying the CpG sites that most significantly differed by tissue status. They evaluated the 254 CpG sites on the array that cover 110 miRNAs from 64 host genes, and detected aberrant miRNA gene methylation in hepatocellular carcinoma tissues (10).

In the current study, we tested the hypothesis that DNA methylation alterations in CpGs associated with miRNA encoding genes may be one of the causes of a deregulated mature miRNA expression in relation to colorectal cancer (CRC) and breast cancer (BC) onset. We described the results of two case-control studies nested in a prospective cohort in which we measured DNA methylation in peripheral white blood cells of subjects who later developed CRC or BC, and subjects who remained clinically healthy. We estimated DNA methylation from the Infinium Human Methylation450 BeadChip (HM450) array, assuming that altered miRNA methylation may be present before the diagnosis and can be associated with cancer risk.

## Material and methods

### Subject recruitment

Study participants were drawn from the Italian component of the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. This subcohort consists of 46857 volunteers, recruited from five different centres within Italy (Varese, Turin, Florence, Naples and Ragusa) with standardized lifestyle and personal history questionnaires, anthropometric data and blood samples collected for DNA extraction (11).

For the present investigation, two nested case-control studies were designed, one on BC and one on CRC. Incident cases were identified through Cancer Registries with <2% losses to follow-up. Controls were individually matched on age ( $\pm 5$  years), sex, seasonality of blood collection and length of follow-up. All participants signed a written informed consent and the ethical review boards of the International Agency for Research on Cancer, and of the collaborating institutions responsible for subject recruitment in each of the EPIC recruitment centres approved the study.

### Anthropometric and dietary/lifestyle factors

Height and weight were measured at enrolment with a standardized protocol and body mass index (BMI calculated as the ratio between weight in kg and squared height in metres) was treated as a continuous variable. Data on smoking status were collected at study enrolment through questionnaires and participants were categorized as never, former and current smokers. Information on dietary and alcohol intake was obtained via a semiquantitative Food Frequency Questionnaire from which estimated consumption of foods and alcohol in g/day was calculated (12). Alcohol consumption was treated as a continuous variable (g/day). Physical activity was assessed using the Cambridge Physical Activity Index (13), which combines self-reported occupational activity with time participating in cycling and other sports. Participants were divided into four categories: inactive (sedentary job and no recreational activity), moderately inactive

(sedentary job with >1h of recreational activity per day), moderately active (standing or physical job with some recreational activity) and active (standing or physical job with some recreational activity or a heavy manual job).

## DNA preparation and Illumina Infinium Human methylation platform

DNA samples were extracted from buffy coats using the QIAasympy DNA Midi Kit (Qiagen, Crawley, UK). Bisulphite conversion of 500 ng of DNA from each sample was performed using the EZ-96 DNA Methylation-Gold™ Kit according to the manufacturer's protocol (Zymo Research, Orange, CA). Bisulphite-converted DNA was used for hybridization on the HM450 array, following the manufacturer's protocol. The array measures DNA methylation at 485 512 cytosine positions across the human genome, of which 482 421 CpG sites and 3091 non-CpG sites; hereafter the term CpG will be used to refer to all of these, unless otherwise specified. A whole genome amplification step was followed by enzymatic end-point fragmentation and hybridization to HM450 arrays at 48°C for 17 h, followed by single nucleotide extension. The incorporated nucleotides were labelled with biotin (ddCTP and ddGTP) and ddNP (ddATP and ddTTP). After the extension step and staining, the BeadChip was washed and scanned using the Illumina HiScanSQ system. Sample quality was assessed using control probes present on the microarrays. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software, which normalizes within-sample data using different internal controls that are present on the HM450 array and internal background probes. The methylation level at each CpG was expressed as a  $\beta$ -value, which represents the fraction of methylated cytosines at that specific location. Data preprocessing was carried out using in-house software written for the R statistical computing environment as described in [14]. Probes not detected ( $P > 0.05$ ) in more than 20% of samples were removed from the dataset. Missing data were first imputed using the *k*-nearest neighbours method as implemented in the R package 'impute' [15]. The whole methylation data for the study population have been deposited and are available online (GSE51057).

## Descriptive analysis of probes associated to miRNAs

In the present study, the analyses were focused on probes associated to an Illumina annotation for genes encoding for miRNAs. In the HM450 platform, 3439 probes are associated with at least one miRNA.

The correspondence between annotations provided by Illumina manifest and the proximal gene occurring within a window of  $\pm 100$  kbp respect to each probe were verified in the preliminary selection of the probes related to miRNA-encoding genes. Only those miRNAs associated to an Illumina annotation also stored in both mirBase (release 20) and HUGO Gene Nomenclature Committee database (<http://www.genenames.org/>) were selected. In the preliminary phase, the following subgroups of probes were discarded: (i) those mapping on chromosome X, (ii) those including single-nucleotide polymorphisms (SNPs) (except for SNPs with minor allele frequency = 0 in Caucasians from 1000 Genomes) and (iii) those associated to discordant miRNA annotations. Some inconsistencies between the probe position and the coordinates of the associated gene were detected considering the transcription start site (TSS) as reference point. More in detail, each probe is usually associated with a gene annotated on a specific chromosome, but in some instances an inconsistency was found verifying the genomic coordinates of the associated gene in another repository. After the resolution of these contradictions (i.e. removing the incorrectly associated probes) and following the workflow described in Figure 1A, a set of 2191 probes (*IlluProbes*, 63.7% of the whole set of miRNA-related probes) has been obtained which is associated with 517 unique miRNA annotations (*IllumiRNAs*, 84.1% of the total miRNAs annotated in HM450 platform and in mirBase) used in the following analysis. Their genomic distribution is reported as Circos plot in Figure 1B. miRNA genomic distribution is essentially reflected in the number of probes in *IlluProbes*, except for chromosome 9 where the number of miRNAs exceeds the amount of probes. Two spikes of probes and miRNAs can be identified in chromosomes 14 and 19, corresponding to two well-known miRNA clusters [16]. To assess probe and miRNA distribution over the genome, a probes/miRNAs ratio was computed. It is possible to observe a non-normal data distribution (Shapiro–Wilk Test,  $P < 0.0001$ ) that reveals

an average number of four probes associated to each miRNA (range 1–16), with 25% of miRNAs having only one probe associated, Figure 1C.

The *IlluProbes* list contains two classes of probes: 'unique', associated only to one miRNA and 'multi', associated to multiple miRNAs. Moreover, miRNAs may be associated to a probe signature composed by: (i) unique probes (distinctive), (ii) miRNA signatures composed by probes belonging to the multiclass (collective) and (iii) miRNAs associated to unique and multi probes (overlapping). Probes and miRNAs counts in each category are reported in Supplementary Table 1, available at Carcinogenesis Online.

The genomic context and CGI distribution of *IlluProbes* were investigated resulting in 89.3% of the probes mapped into promoter regions (within 200 and 1500 bps from a TSS) while the others were mapped in the body region of genes (Supplementary Figure 1A left, available at Carcinogenesis Online). Furthermore, more than 50% of *IlluProbes* fell into regions without CGI (Supplementary Figure 1A right, available at Carcinogenesis Online). Additionally, we verified the genomic annotations of probe and miRNA coordinates based on ENSEMBL (version 75). In particular, we computed the frequencies of four genomic groups: coding gene, non-coding region, intergenic or overlapping between a coding and a non-coding region. Note that non-coding regions include long non-coding RNAs, antisense, pseudo-genes, processed transcripts and sno/micro RNAs as reported in Supplementary Figure 1B, available at Carcinogenesis Online, which reports two pie charts that contain the *IlluProbe* and *IllumiRNA* counts.

## Statistical analysis

The distribution of anthropometric and dietary/lifestyle factors in the study group was compared between cases (both BC and CRC or separately) and controls using Chi-square (categorical data) and Kruskal–Wallis (continuous data) tests.

Considering all probes for each *IllumiRNA* locus, a  $\beta$ -value resulting from the average of all  $\beta$ -values of this probe signature was computed. To identify differentially methylated miRNAs (DMmiRNAs) between two conditions (both for covariates analyses and case–control studies), an unpaired Wilcoxon Rank Sum test with Bonferroni correction was performed. In particular, for each covariant considered (age at recruitment, smoking status, BMI, alcohol consumption, meat and folic acid intake and physical activity index) samples were divided into two categories, according to median for continuous variables. A difference was declared statistically significant if the Bonferroni-adjusted  $P \leq 0.05$ . The possibility that the identified associations could be explained by the confounders was verified by multivariate linear regression analysis.

The DMmiRNAs were used to cluster samples associated to similar methylation profiles, regardless of the case–control status, by the recursively partitioned mixture model (RPMM) implemented in the R package RPMM [17]. Logistic regressions were computed considering either only selected risk factors or both risk factors and previously identified RPMM classes. Subsequently, the area under the curve was computed and the results were compared by the De Long test as described in [18].

To validate the most relevant results from the case–control studies, the dataset TCGA\_BRCA\_hMethyl450 by The Cancer Genome Atlas (TCGA) consortium [19] was downloaded through the UCSC Cancer Genomics Browser (<https://genome-cancer.ucsc.edu/>) [20]. This dataset includes 734 primary BC and 98 normal tissue methylation profiles. For each identified DMmiRNA in the present study, a  $\beta$ -value was computed following the same approach as for our dataset. To identify DMmiRNAs, a Wilcoxon Rank-Sum test with Bonferroni correction was performed between cancer and control tissues.

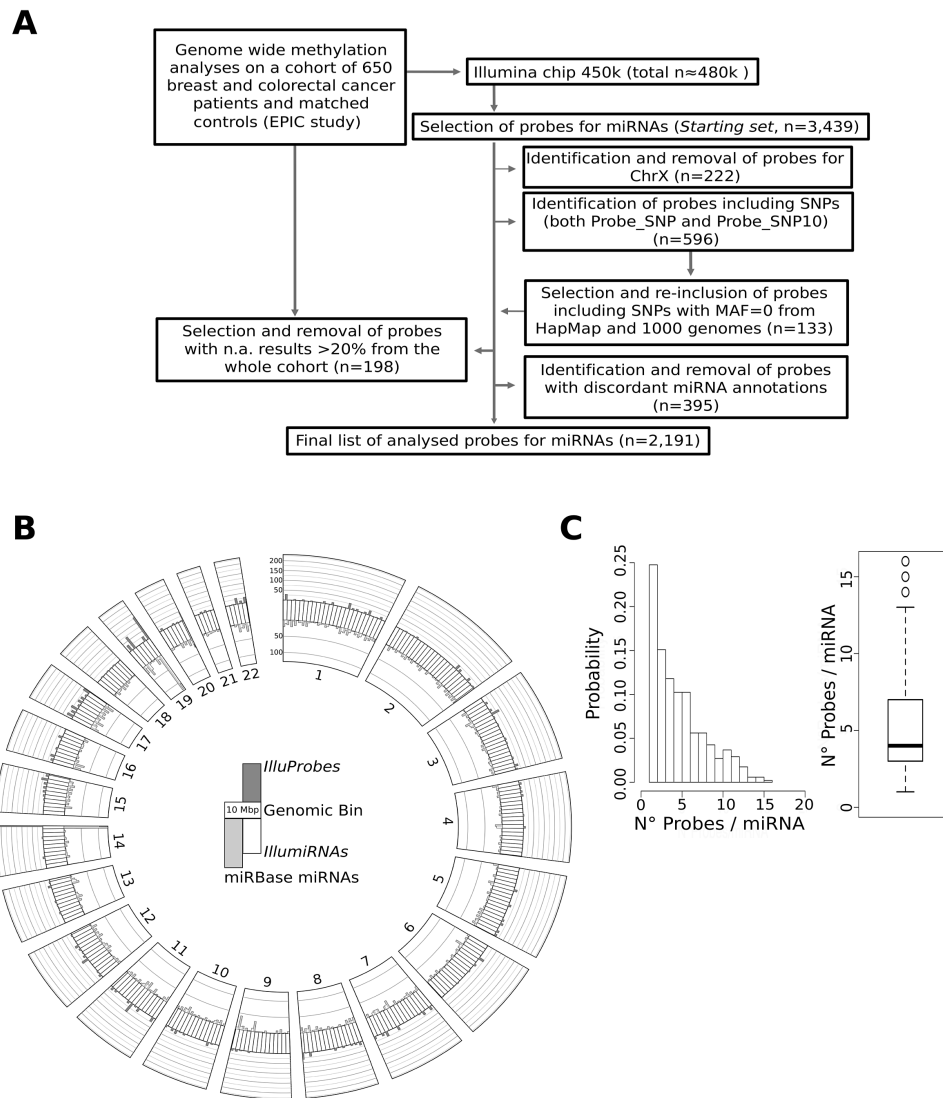
All statistical analyses were conducted using the open source R environment ([www.r-project.org](http://www.r-project.org)).

## Results

### Study population

The study was carried out among 650 individuals: 166 that developed BC, 159 that developed CRC and 325 matched healthy controls. Blood samples from cancer cases were taken on average 64.6 months (range: 0.5 to 187.9) before diagnosis of BC, and





**Figure 1.** (A) Work flow applied in the analysis to extrapolate the probe set. At each step of filtering, the resulting numbers of probes are shown. (B) Circos representation of the genomic distribution of human miRNAs and the Illumina HumanMethylation450K probes residing in miRNA genes. Each ideogram represents a human chromosome split in bins of 10 Mbp. Each bin is illustrated as a white segment delimited by vertical black lines. The number of probes used in the analysis (*IlluProbes*, dark grey bars) is reported on the top, while the number of resulting miRNAs (*IllumiRNAs*, white bars), and those annotated in miRBase (light grey bars) are shown at the bottom. (C) Histogram (left) and box plot (right) of the distribution of the number of probes associated to each miRNA.

74.1 months (range 0.2 to 172.8) for CRC. The main clinical, biological and lifestyle characteristics of the study population are summarized in [Table 1](#). Controls and cases had the same age at recruitment and were matched by sex; moreover, they did not significantly differ for any recorded confounders, except for smoking status among BC cases and controls ( $P = 0.018$ ), and BMI between cases and controls in the whole group ( $P = 0.027$ ), and in the CRC group ( $P = 0.028$ ).

### Methylation profiles analysis

To evaluate the differences in methylation profiles in the present population, we analysed the data following these criteria: (i) all subjects together, independently of case-control status, to investigate the role of each assessed dietary and lifestyle factor on raw miRNA methylation profiles at recruitment, (ii) each case-control study separately, to investigate the potential associations between prediagnostic miRNA methylation profiles and either BC or CRC.

Among all the investigated factors (age at recruitment, smoking status, BMI, alcohol consumption, meat and folic acid

intake and physical activity), four factors were associated to significantly differentially methylated CpG sites of miRNAs after Bonferroni adjustment. In detail, 117 DMmiRNAs were associated with gender (39 hypermethylated and 78 hypomethylated in females versus males), 23 with age (17 hypomethylated and 6 hypermethylated in older versus younger subjects, according to median age), 4 with physical activity status (2 hypermethylated and 2 hypomethylated miRNAs in active versus inactive) and 2 with smoking habit (1 hypermethylated in current smokers versus never smokers and 1 hypomethylated in former smokers versus never smokers). All DMmiRNAs are shown in [Supplementary Figure 2](#), available at [Carcinogenesis Online](#).

With respect to the case-control studies, we analysed BC and CRC cases with their own matched controls separately. No DMmiRNAs were detected in CRC patients in comparison with healthy matched controls, also stratifying for tumor location. On the other hand, in BC samples we identified eight DMmiRNAs (miR-328, miR-675, miR-1307, miR-1286, miR-1275, miR-1910,

**Table 1.** Study population characteristics

Characteristics	Category	All controls (n = 325)	All cases (n = 325)	P value	BC controls (n = 166)	BC cases (n = 166)	P value	CRC controls (n = 159)	CRC cases (n = 159)	P value
Age at recruitment (years)	Mean $\pm$ SD	53.53 $\pm$ 7.01	53.51 $\pm$ 7.05	0.525	52.2 $\pm$ 8.08	52.1 $\pm$ 8.08	0.792	54.2 $\pm$ 7.85	54.2 $\pm$ 7.91	0.934
Gender	Median	55	54		54	54		56	55	
	Min/Max	34/71	35/72		35/70	34/70		35/71	35/72	
	Female	241	241	–	166	166	–	75	75	–
	Male	83	83		0	0		84	84	
Time to diagnosis (years)	Mean $\pm$ SD	NA	5.77 $\pm$ 4.03		NA	5.39 $\pm$ 4.44		NA	6.18 $\pm$ 3.52	
Smoking status	$\leq 1$	NA	29		NA	22		NA	7	
	>1 and $\leq 3$	NA	79		NA	51		NA	28	
	>3 and $\leq 5$	NA	52		NA	19		NA	33	
	>5	NA	165		NA	74		NA	91	
	Non-smokers	159	174		86	108	<b>0.018</b>	73	65	0.617
	Smokers	73	69	0.430	38	31		35	38	
Body mass index (kg/m <sup>2</sup> )	Ex-smokers	86	74		42	24		44	50	
	Missing	7	8		0	2		7	6	
	$\leq 20$	13	10	<b>0.027</b>	8	7	0.537	5	3	<b>0.028</b>
	>20 and $\leq 25$	146	109		75	62		71	47	
Physical activity index	>25 and $\leq 30$	118	149		55	70		62	79	
	>30 and $\leq 35$	49	43		22	21		17	22	
	>35	7	12		5	5		2	7	
	Missing	2	2		1	1		1	1	
	Inactive	86	101	0.606	53	64	0.494	33	37	0.768
	Moderately inactive	140	136		75	67		65	69	
Alcohol consumption (g/day)	Moderately active	57	51		22	22		35	28	
	Active	34	30		16	11		18	19	
	Missing	7	8		0	2		7	6	
	Mean $\pm$ SD	12.98 $\pm$ 45.07	13.13 $\pm$ 45.44	0.915	9.00 $\pm$ 13.29	7.19 $\pm$ 11.58	0.188	17.29 $\pm$ 20.07	19.49 $\pm$ 21.68	0.358
Red meat consumption (g/day)	Mean $\pm$ SD	129.44 $\pm$ 53.23	131.95 $\pm$ 53.48	0.640	122.31 $\pm$ 60.51	119.43 $\pm$ 55.57	0.665	136.67 $\pm$ 73.80	144.46 $\pm$ 70.63	0.347
Folate intake ( $\mu$ g/day)	Mean $\pm$ SD	277.66 $\pm$ 82.93	283.90 $\pm$ 84.79	0.409	275.76 $\pm$ 95.19	274.18 $\pm$ 86.23	0.875	279.73 $\pm$ 85.69	294.32 $\pm$ 110.77	0.199

Significant P values in bold.

miR-24-1 and miR-548a1; Figure 2A), which were significantly hypomethylated in cases in comparison with controls (P values ranging from 0.001 to 0.017 after the Bonferroni correction, Table 2). All DMmiRNAs in the BC group were also confirmed after adjustment for covariates in multivariate linear regression analysis (Supplementary Table 2, available at Carcinogenesis Online). For BC, we had the possibility to stratify cases according to morphology (ductal or lobular carcinoma) and invasive/in situ cancer. Apart from the already identified eight miRNAs, three additional differentially methylated miRNAs appeared for invasive breast carcinoma (miR-200B, miR-525, miR-96) and two more for lobular carcinoma (miR-1180, miR-601). However, the number of observations was reduced in subgroup analyses.

To exclude reverse causation, we removed all cases that developed BC within 1 year from the recruitment ( $n = 22$ ). In this analysis, seven out of the eight identified DMmiRNAs were confirmed, with miR-548a1 losing statistical significance. Interestingly, miR-548a1 was the only miRNA whose different methylation levels between cases and controls increased with decreasing time interval between recruitment/sampling and cancer diagnosis (Supplementary Figure 3, available at Carcinogenesis Online).

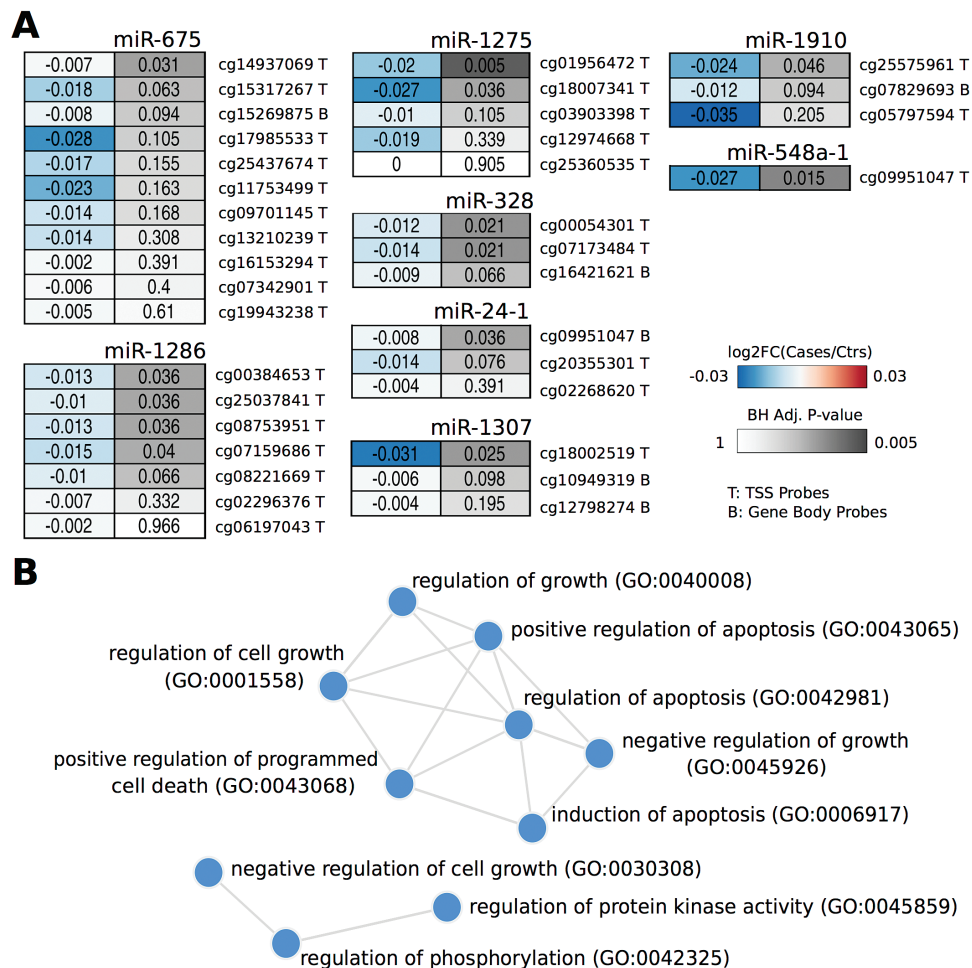
### Discriminatory ability of identified DMmiRNAs

The RPMM method identified seven clusters based on the methylation profiles of the 36 probes associated to the DMmiRNAs

identified among the BC group. Subsequently, two logistic regressions were computed considering: (i) one set of known risk factors for cancer (i.e. physical activity, BMI, alcohol consumption, red meat consumption and folate intake, as reported in Table 1), and those specific for BC (i.e. age at menarche, parity, breastfeeding habits, age at menopause and family history of BC, though not significant as reported in Supplementary Table 3, available at Carcinogenesis Online), and matching variables (i.e. age and EPIC centre of recruitment), namely *model1*, or, (ii) the same set of risk factors/variables and the seven RPMM classes (Figure 3A) identified with respect to the miRNA methylation profiles, namely *model2*. The area under the curve was 0.62 (95% CI: 0.56–0.69) for *model1* and 0.71 (95% CI: 0.65–0.77) for *model2*. Figure 3B shows receiver operating characteristic (ROC) curves of *model1* and *model2*. A statistically significant increment in sensitivity and specificity using both risk factors and epigenomic information was highlighted in *model2* versus *model1* by the De Long test ( $P = 0.003$ ).

### Methylation profiles of DMmiRNAs: validation datasets of TCGA tissue samples

We verified the methylation levels of the eight previously identified DMmiRNAs in 734 primary BC and 98 normal breast tissues collected by the TCGA consortium. Four probes (cg15317267, cg18002519, cg12974668 and cg05797594) out of the 36 associated to the eight DMmiRNAs were excluded since no data were



**Figure 2.** (A) Heat map of log2 fold change methylation differences and the adjusted P value computed for each probe belonging to the 8 DM miRNAs obtained for the BC cohort. The case-control statistical analysis was performed by using Wilcoxon Rank-Sum method. (B) Network representation of Gene Ontology Biological Process terms enriched by Enrichr (22) for the list of validated target genes.

**Table 2.** Differentially methylated miRNAs between subjects who developed BC and matched controls

miRNA	Position	Probes	Locus location	CpG island location	Mean $\beta$ values controls	Mean $\beta$ values cases	log <sub>2</sub> (cases/controls)	Benferroni Adj. P value
miR-24-1	chr9q22.32, C9orf3 intron	3	TSS (2), Body (1)	Island (1), Shore (2)	0.926	0.920	-0.009	1.12 <sup>-02</sup>
miR-328	chr16q22.1, ELMO3 intron	3	TSS (2), Body (1)	Shelf (3)	0.890	0.883	-0.012	9.24 <sup>-04</sup>
miR-548a1	chr6p22.3, intergenic	1	TSS (1)	NA	0.878	0.861	-0.027	1.71 <sup>-02</sup>
miR-675	chr11p15.5, H19 exon	11	TSS (10), Body (1)	Island (5), Shore (6)	0.772	0.765	-0.012	5.87 <sup>-03</sup>
miR-1275	chr6p21.31, intergenic	5	TSS (5)	NA	0.800	0.792	-0.014	7.31 <sup>-03</sup>
miR-1286	chr22q11.21, RTN4R intron	7	TSS (7)	NA	0.868	0.862	-0.01	6.49 <sup>-03</sup>
miR-1307	chr10q24.33, USMG5 intron	3	TSS (1), Body (2)	Shore (3)	0.862	0.855	-0.012	6.00 <sup>-03</sup>
miR-1910	chr16q24.1, C16orf64 intron	3	TSS (2), Body (1)	NA	0.748	0.736	-0.022	1.02 <sup>-02</sup>

For each miRNA, genomic location, overlapped genes, number of probes and their classification are reported. NA, not available.

available. A comparison of methylation levels between tumour and normal tissues showed, in agreement with our findings, a significant hypomethylation for four miRNAs: miR-675, miR-548a1, miR-1910 and miR-1275 ( $P < 0.001$  for all miRNAs, except miR-1275  $P = 0.041$ ) while for miR-1307 hypomethylation was non-significant. In contrast, miR-24-1, miR-328 and miR-1286 resulted hypermethylated in neoplastic tissue compared to control (Supplementary Figure 4, available at Carcinogenesis Online).

### Target genes and gene enrichment analysis

We investigated targets of the identified DMmiRNAs by using the mirWalk database (21), considering only the validated target annotations. Three DMmiRNAs (miR-328, miR-1910 and miR-548a) had Dicer1 as common target, while two DMmiRNAs (miR-328 and miR-675) had the KRAS gene. No other overlap emerged (see Supplementary Table 4, available at Carcinogenesis Online for the complete list of targets). We performed a functional enrichment analysis on the identified validated targets using the Enrichr algorithm (22). GO biological process terms associated with a significant  $P$  ( $P < 0.001$ ) are reported in Figure 2B. Notably, significant terms were mainly related to apoptosis (GO:0006917, induction of apoptosis, adjusted  $P < 0.001$ ) and growth pathways (GO:0030308, negative regulation of cell growth, adjusted  $P < 0.001$ ).

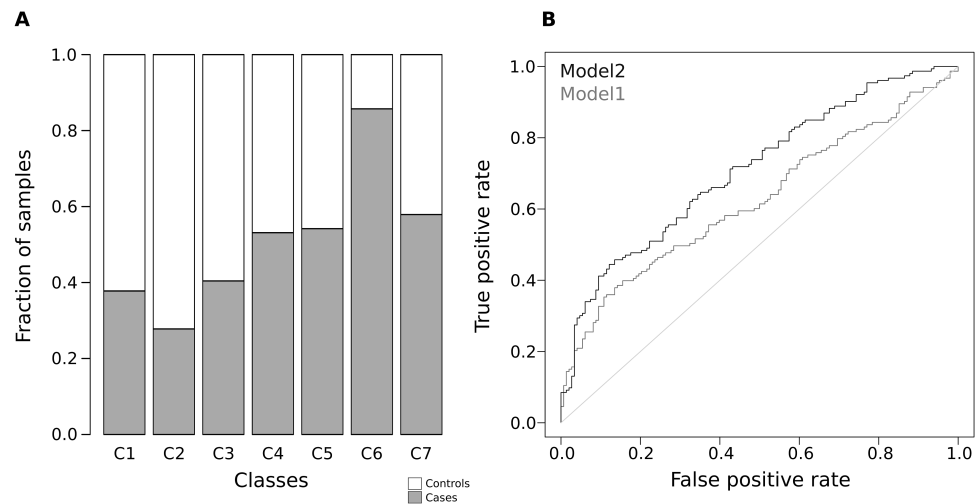
### Discussion

Aberrant DNA methylation of miRNA encoding genes has received attention as an emerging mechanism for miRNA deregulation in cancer (23–25). If the DNA methylation patterns significantly correlate with repression/upregulation of relevant miRNAs, the biological effects should be greater than the effect on a single protein-coding gene since a single miRNA can regulate multiple downstream target genes. These amplified effects indicate that miRNAs may be potentially more sensitive biomarkers to detect early malignant transformation and improve diagnosis and prognosis.

In the present study, we investigated the methylation levels of miRNA encoding genes in blood samples of subjects from two prospective case–control studies nested in the EPIC-Italy cohort. The main finding is that we have identified eight DMmiRNAs in prediagnostic samples of subjects who developed BC during follow-up, in comparison with healthy subjects matched for main potential confounders. Interestingly, all significant DMmiRNAs resulted less methylated in subjects who developed BC and they, independently of other well-known investigated risk factors, contributed to increase sensitivity and specificity of logistic regression models. In contrast, no DMmiRNAs emerged among subjects who developed CRC.

Among the DMmiRNAs, we have found the intergenic miR-548a1, represented only by one TSS probe and coding for miR-548-3p. This miRNA is part of the miR-548 family, whose encoding genes are located across several human chromosomes (26). To the best of our knowledge, there are no published studies reporting an association between this miRNA and BC. More evidence is available for other miRNAs. miR-675 is located in the first exon of the long non-coding RNA H19 gene that is imprinted and maternally expressed. Long non-coding RNAs and internal miRNAs may have versatile roles in multiple biological processes, including tumorigenesis, as potential non-coding RNA regulatory molecules (27). In particular, H19 has been observed several times in association with different cancers, among which BC, but the underlying mechanism of action remains unclear (27). miR-675 is known to target the tumour suppressor retinoblastoma





**Figure 3.** (A) Distribution of BC cases/controls in the seven identified classes of methylation profiles according to RPMM algorithm. (B) Area under the curve for a set of BC risk factors (*model1*) and the same set and the seven RPMM classes according to miRNA methylation profiles (*model2*).

gene and to promote growth and invasion of neoplastic cells (28). A non-overlapping simultaneous involvement of *H19* and miR-675 in gastric cancer has also been recently shown (29). Few data are available about the methylation status of miR-1910 and its relationship to BC or other cancers. Hypermethylation of its promoter has been observed in cell lines but not in breast tumour tissue (30). miR-1910 is intronic with respect to *C16orf74* (*MGC17624*), a gene locus on chromosome 16q24.1 whose function has yet to be characterized. *C16orf74* expression has been associated with *PRSS3* in bladder cancer (31). miR-1275 is part of a recently described signature of miRNAs upregulated in BC tissue of very young women (32). Notably, miR-548a1, miR-675, miR-1910 and miR-1275 resulted significantly hypomethylated also in breast tumour tissue as compared to normal tissue in the TCGA dataset, indicating that the observed alterations in DNA methylation levels might be clinically relevant.

Both miR-328 and miR-1307 were recently related to BC by studying a large cohort of samples from the TCGA consortium, as part of a prognostic miRNA/mRNA signature (33). Promoter methylation of miR-1307 was observed in cell lines but not in tumour tissue (30). MiR-328 regulates breast cancer resistance protein (BCRP) expression in cancer cells: the observed large interindividual differences in various human tissues has been ascribed to DNA methylation in the miR-328 promoter region (34). Finally, miR-24-1, located within an intron of *C9orf3*, forms a cluster together with miR-23b, miR-27b and miR-3074 whose ectopic expression has been implicated in BC progression (35). In our study, miR-24-1 presented one significantly hypomethylated probe in the body of the gene, while only TSS probes were differentially methylated in all other DMmiRNAs. Interestingly, miR-24-1 was also hypermethylated in tumours in comparison with healthy tissues in the TCGA BC dataset. miR-24-1 was previously shown to be increasingly expressed in BC tissue, promoting its development (36). In addition, miR-24-1 was observed as differentially expressed in blood samples of healthy controls versus patients with early stage BC (37,38).

We also investigated the validated target genes of the eight DMmiRNAs, and a functional enrichment analysis highlighted that significant terms were mainly related to apoptosis and growth pathways. In a general scheme, overexpression of miRNAs, due to hypomethylation, might have a role in cancer onset by negatively regulating the above gene pathways. It is also expectable that the effects of such modest differences in

methylation levels, while perhaps of little consequence individually, may globally affect transcription of a biological process or functional network when they co-occur with other changes in the methylome (39).

Epigenetic regulation is often studied in the context of environmental and population health, as DNA methylation patterns are known to be affected by environmental, lifestyle and demographic factors that affect complex disease risk, such as diet, carcinogen exposure, reproductive factors and age, in different tissues including blood (40–42). Several lines of evidence are accumulating on differential miRNA expression, detected also in blood samples, according to age, sex, BMI, smoking activity and diet (43–45). When investigating potential confounders, the main differences in our study population were observed for gender: approximately two-thirds of miRNA encoding genes were hypomethylated in females. We also observed several DMmiRNAs in relation with age, with a predominance of hypomethylated miRNAs in older subjects. Zongli and Taylor (46) recently reported strong differences between island and non-island sites for the direction of methylation changes with age: while 80% of CpGs at island sites were increasingly methylated with age, 95% of CpGs at non-island sites became progressively demethylated. From our dataset, the majority of *Illu* probes are not located on CGI, in line with above findings. Fewer DMmiRNAs were found associated to physical activity levels and smoking habits, while no association was observed for BMI and dietary factors including alcohol consumption.

We are aware of some limitations of the current study. Despite the genome-wide approach to methylation analysis provided by the HM450 array and the relatively vast number of miRNA encoding genes represented (>600), the list is not fully exhaustive for the human genome, thus resulting more a candidate approach. Moreover, many of the initially available probes were removed for the presence of SNPs, for being on chromosome X, for not being correctly annotated or not being consistently analysed in a large proportion of the subjects (>20%). Another important aspect is that methylation status was investigated in blood cells which are not the target tumour tissue. Several studies have assessed methylation in this surrogate tissue in relation to cancer (47) and in particular BC (48). Also for miRNA expression levels some relationships have been observed (38), particularly in normal tissue before disease onset (as reviewed in 42). However, in blood cells altered methylation levels may be

unrelated to those in the primary tumour tissue, especially in earlier stages of the disease. In solving the question of whether DNA methylation at a specific locus actually influences transcriptional activity, researchers should also aim to establish whether the small DNA methylation differences often observed between groups in surrogate tissues translate into differences in gene expression in the relevant tissue. In this context, we have queried the data generated on tissue samples by perusing the TCGA database, where the expression levels of four DMmiRNAs were available in the BC dataset: the only significant DMmiRNA was miR-24-1, whose expression was lower in tumour tissue, in line with hypermethylation in tumour samples. Finally, an important issue is that methylation levels in whole blood may be biased due to inter-individual heterogeneity of leukocyte subsets (49). In light of the considerations of the work of Reinius *et al.*, we additionally investigated whether any of the *IlluProbes* in our set of DMmiRNAs fell among the list of CGI differentially methylated between different blood cell populations. Only four probes (two for miR-1910 and one, respectively, for miR-328 and miR-675) were identified among those potentially having cell type specific methylation status. However, none of them were among the list of probes used for the estimation of cell type proportion in the Houseman algorithm (50) and, in addition, the use of average values for each of these *IllumiRNAs* in the statistical analyses reduce the potential bias due to the possible different cell composition.

Among the strengths of the present study, we can include the prospective nature of our investigation. Longitudinal studies are invaluable for establishing the temporal sequence and stability of disease-associated epigenetic variations, and hence help distinguish causal (driver) from consequential (passenger) epigenetic changes (47,51). In our case, to take care of the reverse causation phenomenon, we repeated the analyses excluding subjects who developed cancer within 1 year from the sampling; all previously identified DMmiRNAs remained significant, with the only exception of miR-548a1. Interestingly, miR-548a1 was the only DMmiRNA showing a clear trend of decreasing methylation levels in relation to the time elapsed between recruitment and diagnosis. This finding suggests a possible role of miR-548a1 as an early marker for BC. For the analysis of miRNA methylation status, we have adopted a miRNA-centred approach, rather than the widely used probe-centred analysis. An approach similar to ours was employed very recently by Aure *et al.* (2) in BC tissue samples, where epigenetic (methylation) and genetic (copy number) markers were investigated in respect to miRNA expression. Despite the fact that we made use of a bioinformatics approach in understanding miRNA epigenetic status, all statistical analyses were also performed at the individual probe level. Thirteen out of 48 significantly differentially methylated probes were associated to the eight DMmiRNAs previously identified in the BC case-control study.

Present results indicate that a panel of hypomethylated miRNAs measured in prediagnostic blood samples may be associated with BC risk. In this regard, more population-based genome-scale investigations are needed to document interindividual differences in DNA methylation and gene expression via miRNA regulation to evaluate the usefulness of an epigenetic cancer risk marker.

## Supplementary material

Supplementary Tables 1–4 and Figures 1–4 can be found at <http://carcin.oxfordjournals.org/>

## Funding

Compagnia di San Paolo, Torino, Italy (to P.V.); Associazione Italiana per la Ricerca sul Cancro (AIRC, IG 2012, project no: 13236) Italy (to D.P.).

## Acknowledgements

The authors would like to thank Eric Van Emburgh for his excellent technical support.

Conflict of Interest Statement: None declared.

## References

1. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, 13, 484–492.
2. Aure, M.R. *et al.* (2013) Individual and combined effects of DNA methylation and copy number alterations on miRNA expression in breast tumors. *Genome Biol.*, 14, R126.
3. Baylin, S.B. *et al.* (2011) A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer*, 11, 726–734.
4. Guo, H. *et al.* (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466, 835–840.
5. Lopez-Serra, P. *et al.* (2012) DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene*, 31, 1609–1622.
6. Sato, F. *et al.* (2011) MicroRNAs and epigenetics. *FEBS J.*, 278, 1598–1609.
7. Adams, B.D. *et al.* (2014) Aberrant regulation and function of microRNAs in cancer. *Curr. Biol.*, 24, R762–R776.
8. Vrba, L. *et al.* (2011) Epigenetic regulation of normal human mammary cell type-specific miRNAs. *Genome Res.*, 21, 2026–2037.
9. Kita, Y. *et al.* (2014) Epigenetically regulated microRNAs and their prospect in cancer diagnosis. *Expert Rev. Mol. Diagn.*, 14, 673–683.
10. Shen, J. *et al.* (2012) Genome-wide aberrant DNA methylation of microRNA host genes in hepatocellular carcinoma. *Epigenetics*, 7, 1230–1237.
11. Riboli, E. (2001) The European Prospective Investigation into Cancer and Nutrition (EPIC): plans and progress. *J. Nutr.*, 131, 170S–175S.
12. Agnoli, C. *et al.* (2011) A priori-defined dietary patterns are associated with reduced risk of stroke in a large Italian cohort. *J. Nutr.*, 141, 1552–1558.
13. Wareham, N.J. *et al.* (2003) Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr.*, 6, 407–413.
14. Campanella, G. *et al.* (2014) Epigenetic signatures of internal migration in Italy. *Int. J. Epidemiol.* pii: dyu198 (epub ahead of print).
15. Hastie, T. *et al.* (2014) Impute: Imputation for microarray data. R package version 1.42.0.
16. Noguer-Dance, M. *et al.* (2010) The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum. Mol. Genet.*, 19, 3566–3582.
17. Koestler, D.C. *et al.* (2013) Recursively partitioned mixture model clustering of DNA methylation data using biologically informed correlation structures. *Stat. Appl. Genet. Mol. Biol.*, 12, 225–240.
18. Fiorito, G. *et al.* (2014) B-vitamins intake, DNA-methylation of One Carbon Metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. *Nutr. Metab. Cardiovasc. Dis.*, 24, 483–488.
19. Weinstein, J.N. *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, 45, 1113–1120.
20. Zhu, J. *et al.* (2009) The UCSC Cancer Genomics Browser. *Nat. Methods*, 6, 239–240.
21. Dweep, H. *et al.* (2014) miRWalk database for miRNA-target interactions. *Methods Mol. Biol.*, 1182, 289–305.
22. Chen, E.Y. *et al.* (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14, 128.
23. Suzuki, H. *et al.* (2012) DNA methylation and microRNA dysregulation in cancer. *Mol. Oncol.*, 6, 567–578.
24. Radpour, R. *et al.* (2011) Integrated epigenetics of human breast cancer: synoptic investigation of targeted genes, microRNAs and proteins upon demethylation treatment. *PLoS One*, 6, e27355.

25. Lehmann, U. et al. (2008) Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J. Pathol.*, 214, 17–24.
26. Liang, T. et al. (2012) Genome-wide analysis of mir-548 gene family reveals evolutionary and functional implications. *J. Biomed. Biotechnol.*, 2012, 679563.
27. Guo, L. et al. (2014) An integrated evolutionary analysis of miRNA-lncRNA in mammals. *Mol. Biol. Rep.*, 41, 201–207.
28. Shi, Y. et al. (2014) Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675. *PLoS One*, 9, e86295.
29. Li, H. et al. (2014) Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget*, 5, 2318–2329.
30. Vrba, L. et al. (2013) miRNA gene promoters are frequent targets of aberrant DNA methylation in human breast cancer. *PLoS One*, 8, e54398.
31. Kim, W.T. et al. (2010) Identification of C16orf74 as a marker of progression in primary non-muscle invasive bladder cancer. *PLoS One*, 5, e15260.
32. Pena-Chilet, M. et al. (2014) MicroRNA profile in very young women with breast cancer. *BMC Cancer*, 14, 529.
33. Volinia, S. et al. (2013) Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. *Proc. Natl. Acad. Sci. USA*, 110, 7413–7417.
34. Saikumar, J. et al. (2012) Expression, circulation, and excretion profile of microRNA-21, -155, and -18a following acute kidney injury. *Toxicol. Sci.*, 129, 256–267.
35. Ell, B. et al. (2014) The MicroRNA-23b/27b/24 Cluster Promotes Breast Cancer Lung Metastasis by Targeting Metastasis-suppressive Gene Prosaposin. *J. Biol. Chem.*, 289, 21888–21895.
36. Du, W.W. et al. (2013) MicroRNA miR-24 enhances tumor invasion and metastasis by targeting PTPN9 and PTPRF to promote EGF signaling. *J. Cell Sci.*, 126, 1440–1453.
37. Wu, Q. et al. (2012) Analysis of serum genome-wide microRNAs for breast cancer detection. *Clin. Chim Acta*, 413, 1058–1065.
38. Schrauder, M.G. et al. (2012) Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection. *PLoS One*, 7, e29770.
39. Heijmans, B.T. et al. (2012) Commentary: The seven plagues of epigenetic epidemiology. *Int. J. Epidemiol.*, 41, 74–78.
40. Steegenga, W.T. et al. (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)*, 36, 9648.
41. Nelson, H.H. et al. (2011) Global methylation in exposure biology and translational medical science. *Environ. Health Perspect.*, 119, 1528–1533.
42. Brennan, K. et al. (2012) Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer Prev. Res. (Phila)*, 5, 1345–1357.
43. Tarallo, S. et al. (2014) MicroRNA expression in relation to different dietary habits: a comparison in stool and plasma samples. *Mutagenesis*, 29, 385–391.
44. Ross, S.A. et al. (2011) MicroRNA, nutrition, and cancer prevention. *Adv. Nutr.*, 2, 472–485.
45. Duttagupta, R. et al. (2011) Impact of cellular miRNAs on circulating miRNA biomarker signatures. *PLoS One*, 6, e20769.
46. Xu, Z. et al. (2014) Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis*, 35, 356–364.
47. Langevin, S.M. et al. (2014) Leukocyte-adjusted epigenome-wide association studies of blood from solid tumor patients. *Epigenetics*, 9, 884–895.
48. Barrow, T.M. et al. (2014) Epigenetic epidemiology of cancer. *Biochem. Biophys. Res. Commun.*, 455, 70–83.
49. Reinius, L.E. et al. (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*, 7, e41361.
50. Houseman, E.A. et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13, 86.
51. Rakyan, V.K. et al. (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, 12, 529–541.